Extra cases Setting 1: Single effects on endpoint of interest

After reviewing the results in Tamvada (2023), we decided to assess performance across settings with varying numbers of covariates and different proportions of true covariates. Specifically, this involved using an intermediate sample size ($p = 500$) and changing the number of true covariates. While the previous analysis kept the number of covariates constant, in the new setup we maintained the same ratio of variables to ensure consistency.

1 Simulations

Each simulation was run 20 times with a sample size of $n = 400$. The covariate data, X, for the p predictors were generated from a multivariate normal distribution with a mean of zero, $X \sim \mathcal{N}(0, \Sigma)$. The covariance matrix Σ was defined to have an exchangeable correlation structure where the correlation between the true influential covariates was $\rho = 0.5$. The simulations were designed for a non-proportional hazards setting. For a scenario with p total predictors and k true predictors (e.g., $k = 20$), the coefficient vectors for the two events (β_1 and β_2) were defined as: In this structure, $k/2$ predictors influence event 1, and a distinct set of $k/2$ predictors influences the competing event, with all other $p - k$ predictors having no effect.

The competing models for the experiment are the following:

- **Elastic-net casebase (enet-casebase):** The penalty parameter, λ_{min}, is tuned using 5-fold cross-validation due to computational constraints. To fit a Weibull hazard, the time variable is included as a covariate in the form $\log(t)$. This model is trained with SVRG.

- **Elastic-net casebase (enet-casebase-Acc):** This model is also trained with AccSVRG.

- **Elastic-net Independent Cox (enet-iCR):** For this model, the penalty parameter, λ_{min}, is tuned using 10-fold cross-validation, with the partial likelihood deviance used as the performance metric.

- **Elastic-net Cox with shared penalty (enet-penCR):** This model's penalty parameter, λ_{min}, is tuned using 10-fold cross-validation across a 30 ×30 grid. The Brier score for event 1 serves as the performance metric.

1.1 Results

The following plot shows the variable selection metrics such as Sensitivity, Specificity, and MCC.

MSE of Coefficients

Model ▭ enet-casebase-Acc ▭ enet-iCR ▭ enet-penCR ▭ postenet-casebase-Acc
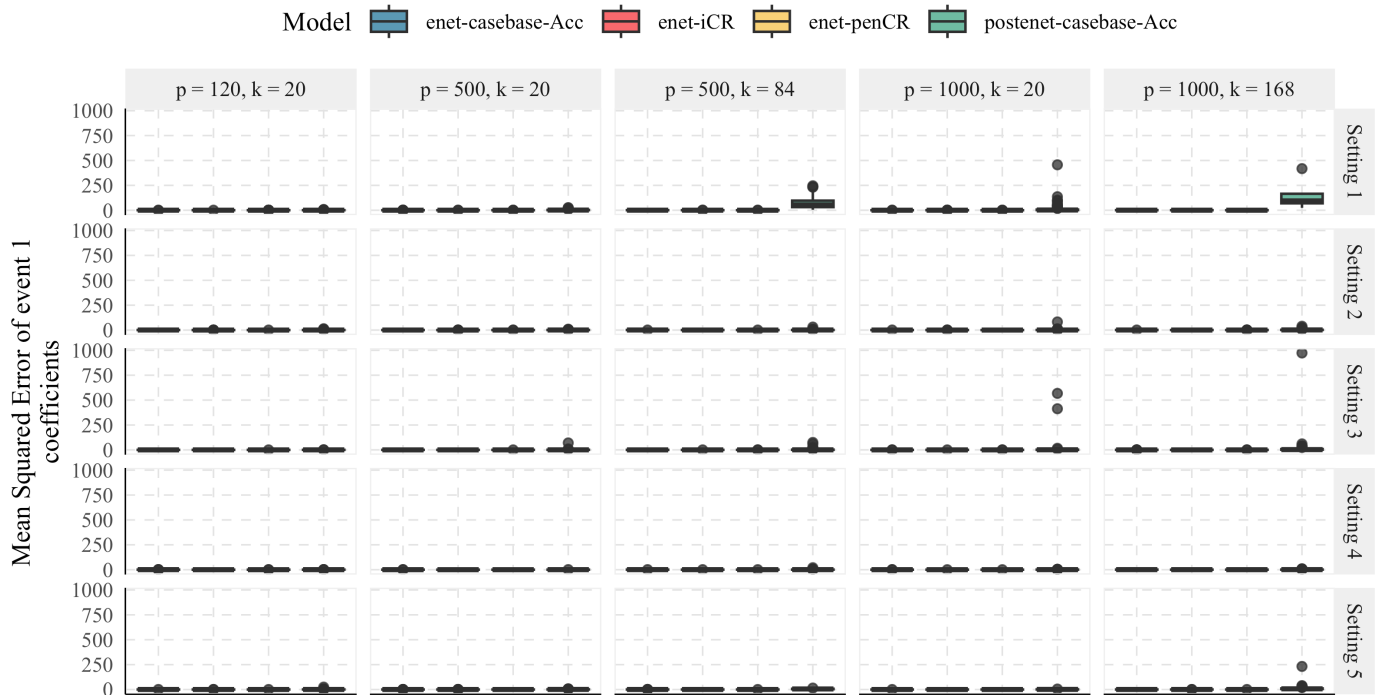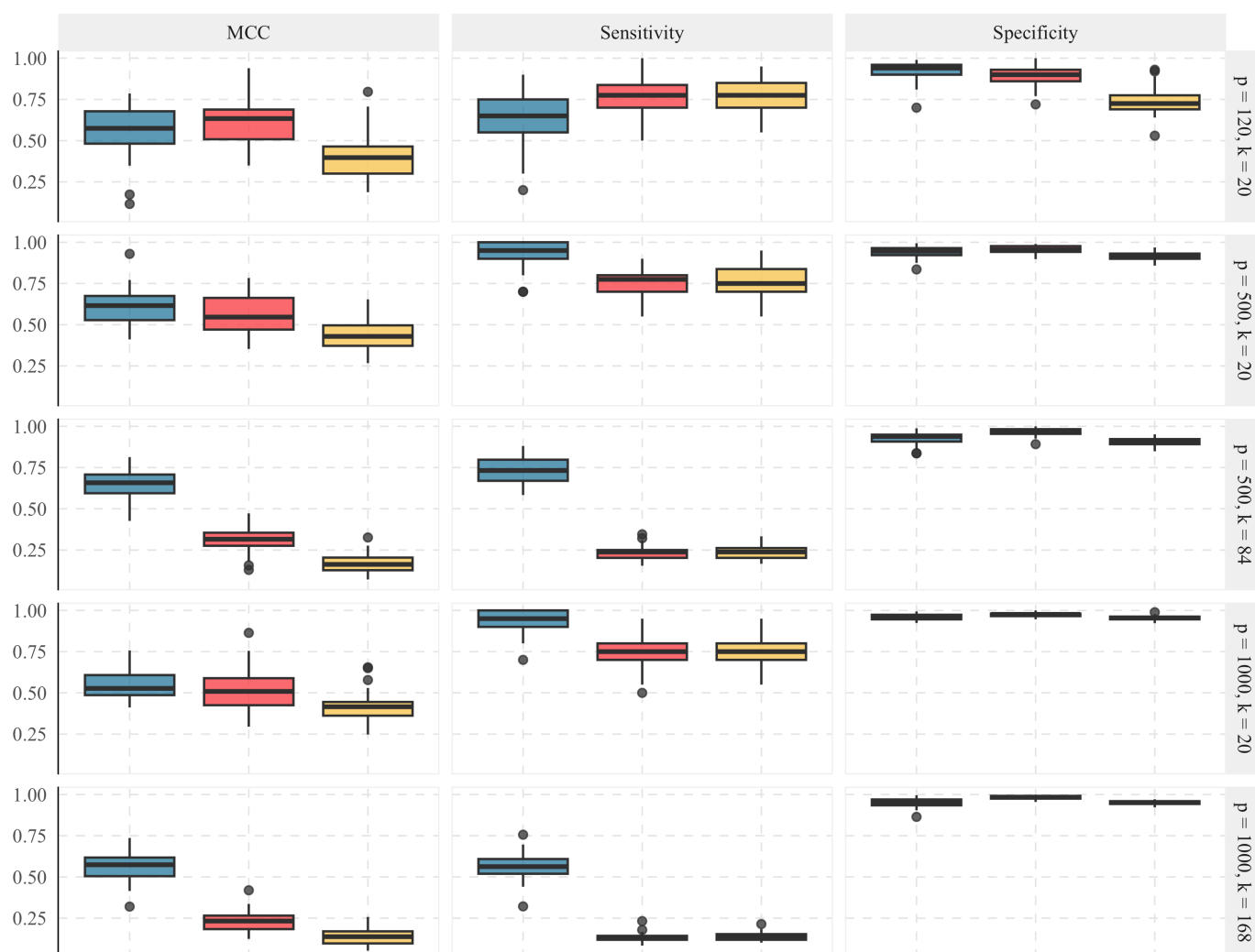


```
coefs_mod %>%
  filter(sim <= 50,
         setting == 1) %>%
  group_by(name, model, setting, p, k, dim) %>%
  summarise(selec_mea = vec_varsel_perc(coef, p, k, setting)) %>%
  unnest(selec_mea) %>%
  arrange(setting, p, k) %>%
  filter(!str_detect(model, "post|bias")) %>%
  pivot_longer(Sensitivity:MCC,
            names_to = "metric",
            values_to = "value") %>%
  ggplot(aes(x = model, y = value,
         fill = model)) +
  geom_boxplot(show.legend = T) +
  facet_grid(fct_inorder(dim)~metric) +
  theme(axis.text.x = #element_text(angle = 90, vjust = 0.5, hjust=1)
         element_blank()) +
  labs(title = "Selection Performance",
      subtitle = "Setting 1",
      fill = "Model",
      y = element_blank(),
      x = element_blank())
```

Selection Performance
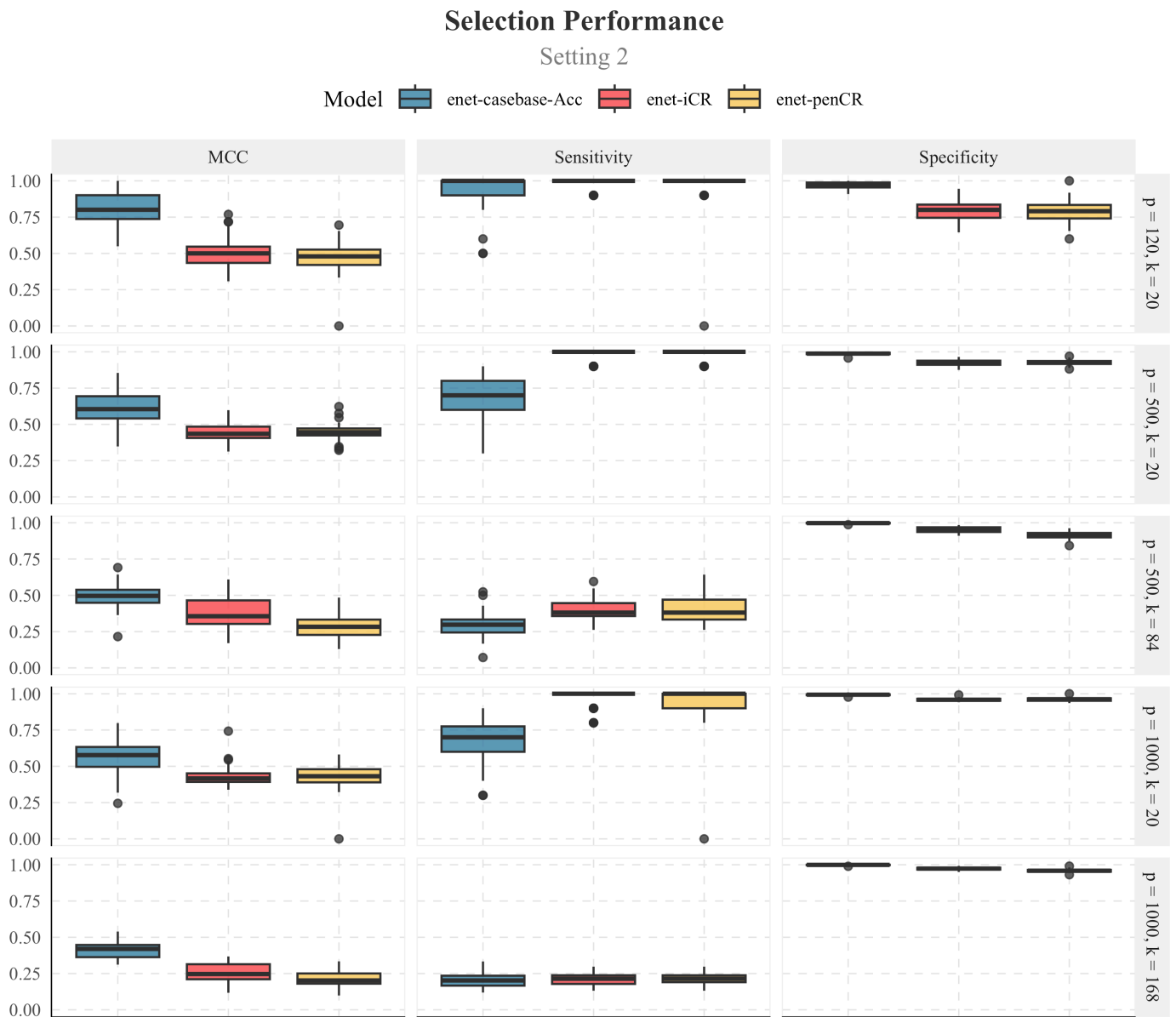
Setting 1


```
coefs_mod %>%
  filter(sim <= 50,
         setting == 2) %>%
  group_by(name, model, setting, p, k, dim) %>%
  summarise(selec_mea = vec_varsel_perc(coef, p, k, setting)) %>%
  unnest(selec_mea) %>%
  arrange(setting, p, k) %>%
  filter(!str_detect(model, "post|bias")) %>%
  pivot_longer(Sensitivity:MCC,
               names_to = "metric",
               values_to = "value") %>%
  ggplot(aes(x = model, y = value,
             fill = model)) +
  geom_boxplot(show.legend = T) +
  facet_grid(fct_inorder(dim)~metric) +
  theme(axis.text.x = #element_text(angle = 90, vjust = 0.5, hjust=1)
        element_blank()) +
```

```
labs(title = "Selection Performance",
     subtitle = "Setting 2",
     fill = "Model",
     y = element_blank(),
     x = element_blank())
```


Selection Performance

Setting 2

coefs_mod %>%
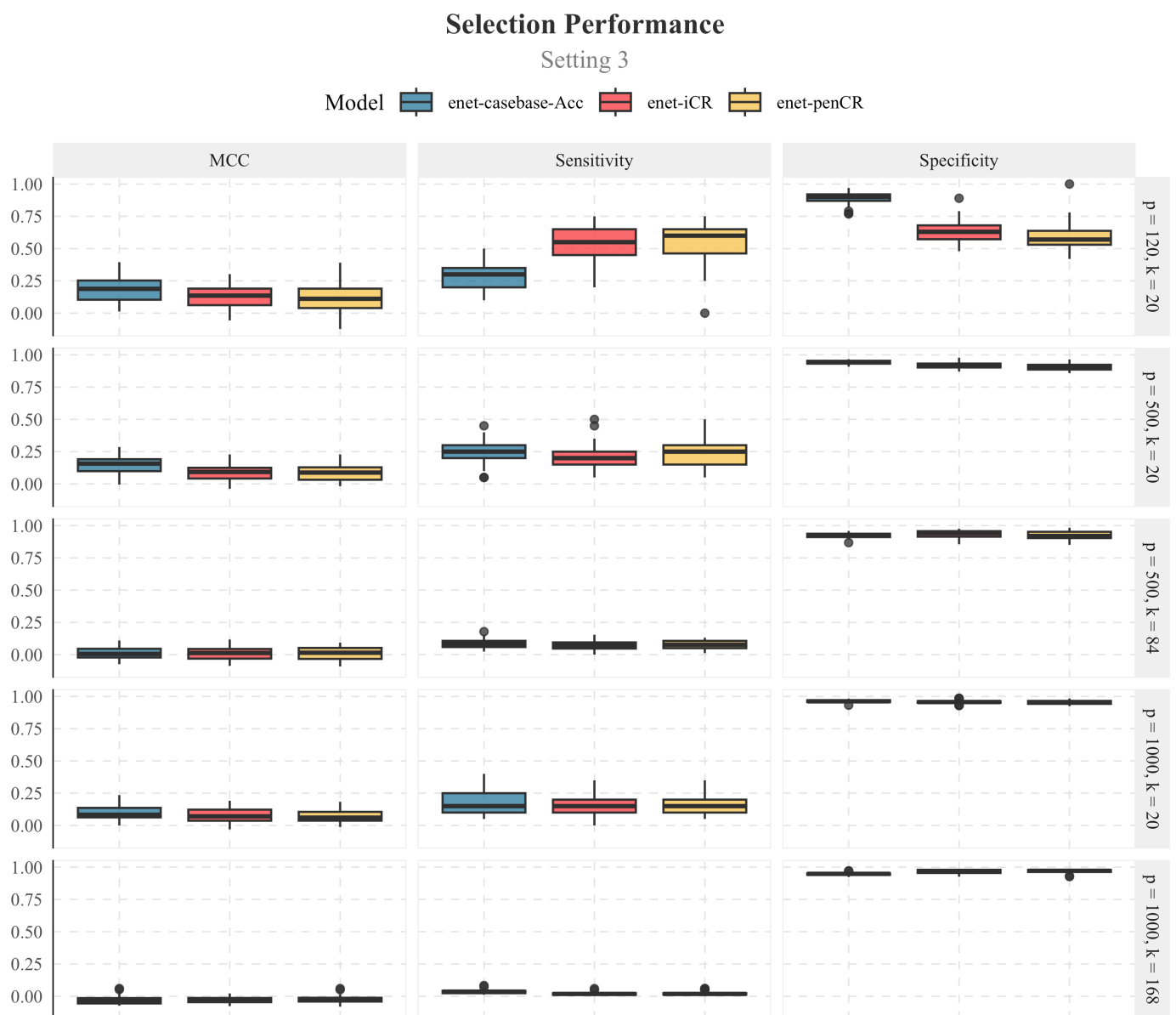```
coefs_mod %>%
  filter(sim <= 50,
         setting == 3) %>%
  group_by(name, model, setting, p, k, dim) %>%
  summarise(selec_mea = vec_varsel_perc(coef, p, k, setting)) %>%
  unnest(selec_mea) %>%
  arrange(setting, p, k) %>%
  filter(!str_detect(model, "post|bias")) %>%
  pivot_longer(Sensitivity:MCC,
               names_to = "metric",
```

```
            values_to = "value") %>%
  ggplot(aes(x = model, y = value,
         fill = model)) +
  geom_boxplot(show.legend = T) +
  facet_grid(fct_inorder(dim)~metric) +
  theme(axis.text.x = #element_text(angle = 90, vjust = 0.5, hjust=1)
         element_blank()) +
  labs(title = "Selection Performance",
     subtitle = "Setting 3",
     fill = "Model",
     y = element_blank(),
     x = element_blank())
```

Selection Performance
Setting 3


```
coefs_mod %>%
  filter(sim <= 50,
       setting == 4) %>%
```

```r
  group_by(name, model, setting, p, k, dim) %>%
  summarise(selec_mea = vec_varsel_perc(coef, p, k, setting)) %>%
  unnest(selec_mea) %>%
  arrange(setting, p, k) %>%
  filter(!str_detect(model, "post|bias")) %>%
  pivot_longer(Sensitivity:MCC,
               names_to = "metric",
               values_to = "value") %>%
  ggplot(aes(x = model, y = value,
             fill = model)) +
  geom_boxplot(show.legend = T) +
  facet_grid(fct_inorder(dim)~metric) +
  theme(axis.text.x = #element_text(angle = 90, vjust = 0.5, hjust=1)
        element_blank()) +
  labs(title = "Selection Performance",
       subtitle = "Setting 4",
       fill = "Model",
       y = element_blank(),
       x = element_blank())
```

```
#> Warning: There were 1600 warnings in `summarise()`.
#> The first warning was:
#> i In argument: `selec_mea = vec_varsel_perc(coef, p, k, setting)`.
#> i In group 5: `name = "setting-4_iter-10_p-1000_k-20"`, `model =
#>   "enet-casebase-Acc"`, `setting = 4`, `p = 1000`, `k = 20`, `dim = "p = 1000,
#>   k = 20"`.
#> Caused by warning in `beta1[nu_ind] <- c(rep(pars_sett1[1], k / 4), rep(c(pars_sett1[2], -
pars_sett1[
#>   2]), k / 8), rep(pars_sett1[3], k / 4), rep(pars_sett1[4], k / 4))`:
#> ! number of items to replace is not a multiple of replacement length
#> i Run `dplyr::last_dplyr_warnings()` to see the 1599 remaining warnings.
```
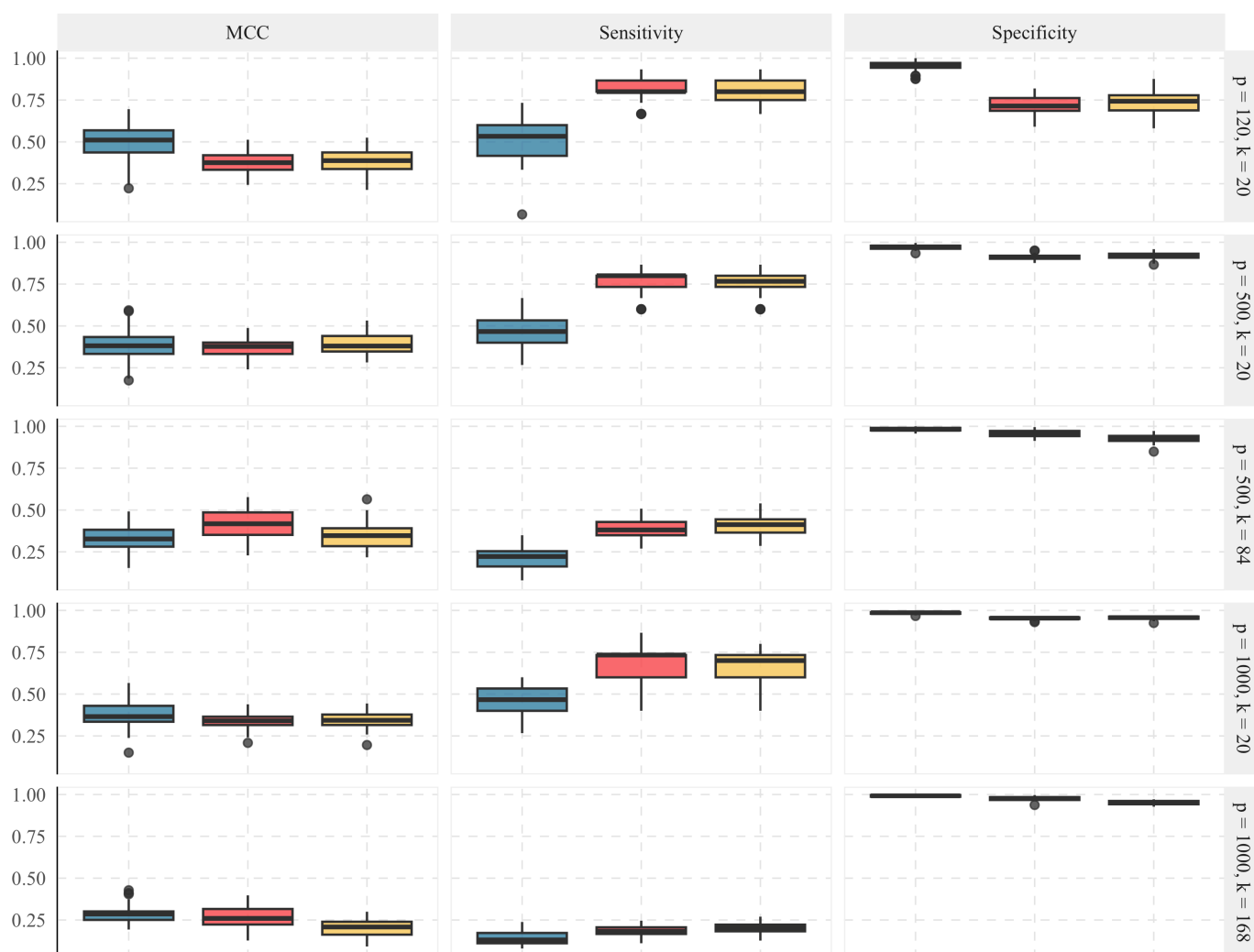
Selection Performance

Setting 4

Model   ▭ enet-casebase-Acc   ▭ enet-iCR   ▭ enet-penCR


```
coefs_mod %>%
  filter(sim <= 50,
         setting == 5) %>%
  group_by(name, model, setting, p, k, dim) %>%
  summarise(selec_mea = vec_varsel_perc(coef, p, k, setting)) %>%
  unnest(selec_mea) %>%
  arrange(setting, p, k) %>%
  filter(!str_detect(model, "post|bias")) %>%
  pivot_longer(Sensitivity:MCC,
               names_to = "metric",
               values_to = "value") %>%
  ggplot(aes(x = model, y = value,
             fill = model)) +
  geom_boxplot(show.legend = T) +
  facet_grid(fct_inorder(dim)~metric) +
  theme(axis.text.x = #element_text(angle = 90, vjust = 0.5, hjust=1)
        element_blank()) +
```

```r
    labs(title = "Selection Performance",
        subtitle = "Setting 5",
        fill = "Model",
        y = element_blank(),
        x = element_blank())
```

```
#> Warning: There were 2096 warnings in `summarise()`.
#> The first warning was:
#> i In argument: `selec_mea = vec_varsel_perc(coef, p, k, setting)`.
#> i In group 1: `name = "setting-5_iter-10_p-1000_k-166"`, `model = "CB-bias"`,
#>   `setting = 5`, `p = 1000`, `k = 166`, `dim = "p = 1000, k = 166"`.
#> Caused by warning in `beta1[nu_ind] <- c(rep(pars_sett1[1], k / 4), rep(pars_sett1[2], k /
#>    pars_sett1[3], k / 4), rep(pars_sett1[4], k / 4))`:
#> ! number of items to replace is not a multiple of replacement length
#> i Run `dplyr::last_dplyr_warnings()` to see the 2095 remaining warnings.
```
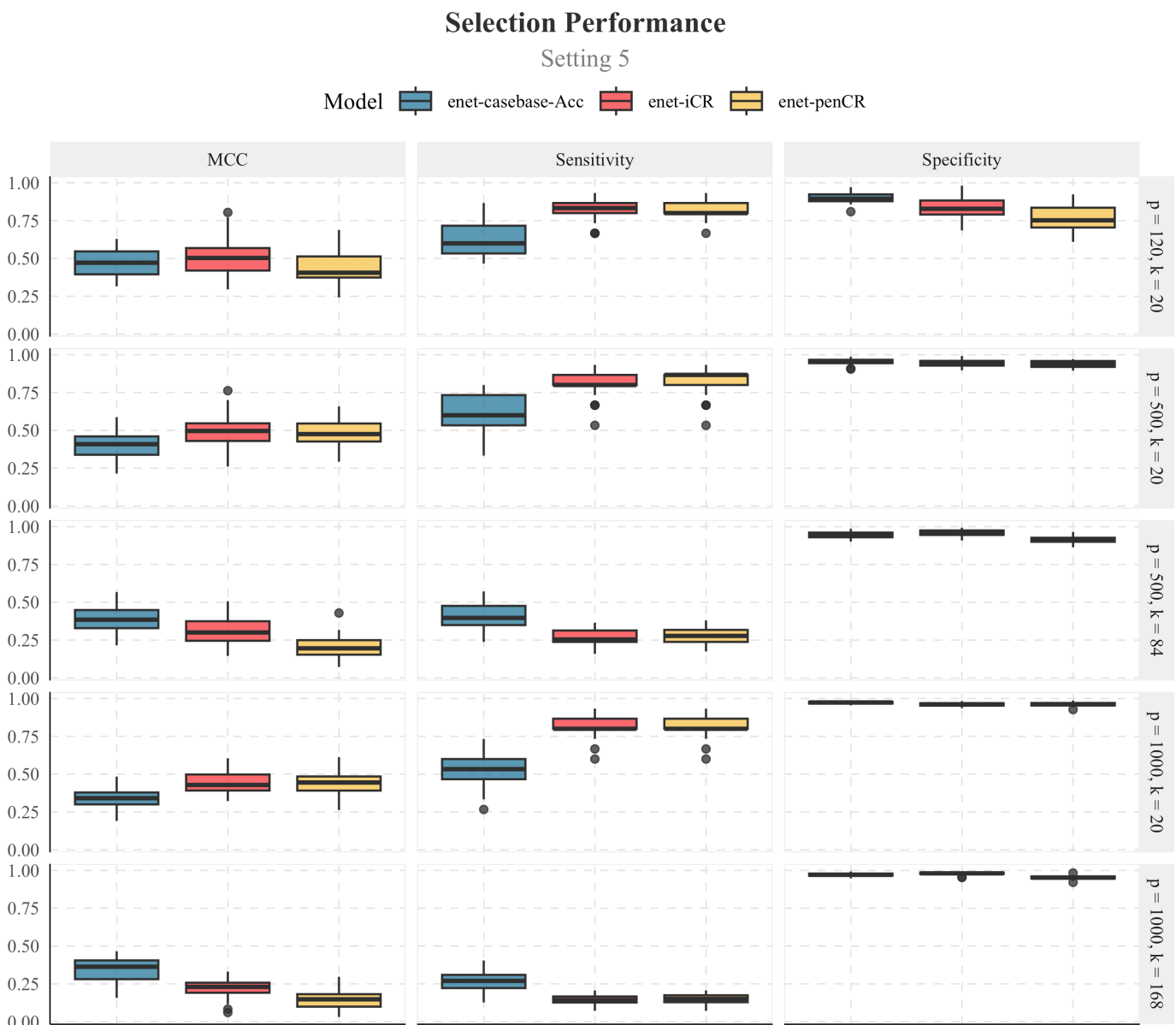

Selection Performance
Setting 5

References

Tamvada, Nirupama. 2023. "Penalized Competing Risks Analysis Using Casebase Sampling."
University of British Columbia. https://doi.org/10.14288/1.0435526.